DOCUMENT RESUME

ED 250 388 TM 840 739

AUTHOR Bolin, Frances
TITLE The Effective Teacher: Educational Testing and Measurement as a Curriculum Concern.
PUB DATE Apr 84
NOTE 30p.; Paper presented at the Annual Meeting of the American Educational Research Association (68th, New Orleans, LA, April 23-27, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Information Analyses (070)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Achievement; *Achievement Gains; Educational Quality; Educational Testing; Efficiency; Elementary Secondary Education; *Evaluation Criteria; Performance Factors; Standards; *Teacher Effectiveness; *Teacher Evaluation; Teacher Supervision
IDENTIFIERS *Barr (A S)

ABSTRACT

The purpose of this paper is to trace the development of the idea of measurement of pupil growth and achievement as a criterion of teacher effectiveness through the scholarly writings of A. S. Barr (1892-1962). The search for the criterion of teacher efficiency, and later effectiveness, was the professional concern that guided Barr's research for four decades. He directed the Wisconsin Studies of Teacher Effectiveness, which influenced the development of objective-based criteria for teacher effectiveness research. His primary interest was in establishing objective, valid criteria of efficient teaching, and, in doing so, he became involved in the measurement of pupil growth and achievement as a criterion of teacher effectiveness. This paper documents chronologically Barr's lifelong research interests including major achievements and shortcomings in his work. (EGS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED250388

THE EFFECTIVE TEACHER: EDUCATIONAL TESTING AND MEASUREMENT AS A CURRICULUM CONCERN

by

Frances Bolin

Teachers College, Columbia University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

F. S. Bolin

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

TM 840 739

A Paper Presented at AERA
Divisions B and F
New Orleans, April 1984

THE EFFECTIVE TEACHER: EDUCATIONAL TESTING AND MEASUREMENT AS A CURRICULUM CONCERN

Application of the principles of scientific management to curriculum theory and practicum might have been set on the shelf along with other doctrines of social efficiency had it not been for the wedding of curriculum theorizing to educational testing and measurement. Parallel to the development of the curriculum field was the growth of the educational testing and measurement movement. As early as 1918 E. L. Thorndike spoke of measurement of the products of learning by quantifying them, presuming that what existed would exist in some amount.[1] Measurement of the products of learning was of interest to those who sought greater precision in curriculum planning. William A. McCall, Professor of Education at Teachers College, Columbia University, wrote that "There are other things in education besides measurement but they have no value so long as they are dissociated from it."[2]

Whether or not school people actually shared McCall's conviction, use of educational testing and measurement came to be of pragmatical importance to them. Public School people, always subject to open scrutiny, could provide tangible evidence of their effectiveness by measuring educational products.[3] Herbert Kliebard points out how extrapolation of principles of Scientific

---

[1] E.L. Thorndike, "The Nature, Purposes and General Methods of Measurements of Educational Products," in Seventeenth Yearbook of the National Society for the Study of Education, pt. 2 (Bloomington, IL: Public School Publishing Co., 1918), p. 16.

[2] William A. McCall, How to Measure in Education (New York: Macmillan Co., 1923), p. 12.

[3] Raymond Callahan, Education and the Cult of Efficiency (Chicago and London: University of Chicago Press, 1961), p. 111.

management to curriculum planning and implementation had the effect of making educational products out of students, having a profound and continuing effect on the field of curriculum itself.[4]

Application of the principles of scientific management, borrowed from business and industry in the early 1900s, to education has been the subject of considerable criticism in recent decades. Raymond Callahan notes that there have been, all along, "strong, intelligent, prophetic voices" of dissent.[5] Yet the modern expression of scientific management, a technological, systems approach to curriculum activity, seems none the weaker for its critics. Putting other criticism aside, however, one might still argue that the usefulness of a technological, systems model in some educational circumstances is not enhanced when it is taken to be the appropriate model in all educational circumstances. Hence it is important to continue to try to understand the history of "scientific" approaches to curriculum theory and practice, and to question why these approaches are so often taken to be the standard in curriculum decision making.

One of the factors contributing to persistence of the technological, systems curriculum model may be that measurement of pupil growth and achievement, which provided the technology for standardization of curriculum outcomes, has been linked to teacher effectiveness. Researchers in the 1920s and 1930s looked for a relationship between teacher efficiency and products of instruction. Following the same course, many researchers of today speak in terms of teacher effectiveness, or competency, pointing to process-product relationships and

---

[4]Herbert M. Kliebard, "Bureaucracy and Curriculum Theory," in Curriculum and Evaluation, ed. Arno A. Bellack and Herbert M. Kliebard (Berkeley: McCutchan, 1977), pp. 612-13.

[5]Callahan, Education and the Cult of Efficiency, p. 125.

links between teacher inputs and product or outcome variables. This phenomenon may have had more than passing influence on the state of curriculum theory and practice today, for when school administrators could point to general or composite measures of pupil growth and achievement as evidence of teacher efficiency, bureaucratic approaches to curriculum decision making were locked into place. School administrators could then speak with authority about system goals, standardization of curriculum inputs and outcomes, and which members of the educational team were serving most effectively, to a public that is always demanding tangible evidence that its tax dollars are well invested, its youth are not compromised in acquiring the "tools for developing their individual powers of mind and spirit to the utmost," and its national security will be preserved.[6]

The purpose of this paper is to trace the development of the idea of measurement of pupil growth and achievement as the criteria of teacher effectiveness through the scholarly writings of A.S. Barr (1892-1962). Searching for the criterion of teacher efficiency, and later effectiveness, was the professional concern that guided Barr's research for four decades. He directed the Wisconsin Studies of Teacher Effectiveness, which influenced development of objective-based criteria for teacher effectiveness research.[7] He was Professor of Education at the University of Wisconsin, Madison, a president of the American Educational Research Association, author and co-author of an influential series of textbooks

[6] The National Commission of Excellence in Education, A Nation at Risk: The Imperative for Educational Reform (Washington, D.C.: U.S. Government Printing Office, 1983), p. 8.

[7] See Arthur M. Cohen, James W. Trent, and Clare Rose, "Research on Teaching in Higher Education," in Second Handbook of Research on Teaching, ed. R.M.W. Travers (Chicago: Rand McNally College Publishing Co., 1973), pp. 997-1171.

on supervision of instruction--the Appleton Series, an editor of the Journal of Educational Research and the Journal of Experimental Education. I will discuss how Barr became interested in finding the criterion of teacher efficiency, where this led in his own research, and how his work has been used in a way that has, I believe, strengthened bureaucratic approaches to curriculum decision making.

Examining this, and other curriculum issues, from an historical perspective continues to be of importance if we are to understand the contextual background of curriculum theory and practice today. An historical study that focuses on one individual's work, however, risks both narrowness and glorification of the individual. From what I have learned about A.S. Barr in my studies he would have wished neither. It is not my purpose either to accuse or exonerate Barr, but to look carefully at how his work developed, how his idea of the efficient teacher became equated with the effective teacher, and how this may have contributed to the present enthusiasm for objective-based teacher performance criteria, which I see as ultimately a curriculum concern.

The Efficient Teacher

How do you know a good teacher when you see one? This question was a continuing source of frustration to A.S. Barr. His search for criteria of effective teaching began as a search for the criterion of efficient teaching. At the time Barr was completing graduate studies at the University of Chicago and at the University of Wisconsin, a great deal of attention was directed toward the application of the principles of scientifical management to curriculum theorizing. Franklin Bobbitt had written How to Make a Curriculum, W.W. Charters was working to reorganize women's education at Steven's College, so that women

could better learn how to be women, and David Snedden was working to standardize objectives of the curriculum theorists. He was classroom teacher and school supervisor turned college professor. Yet his efforts to discover the objective criteria for measurement of teaching efficiency had a subsidiary influence on the shape of curriculum theory and practice today.

Thought, Labor, and Time Saving Standards

Barr was troubled by what he believed to be the typical answer to questions about what constituted good teaching, "Probably when the teacher is attractive, the pupils orderly, and the parents satisfied, the teaching is acceptable."[8] The situation was all the more aggravating because it was widely held, at least in the professional literature, that appearances were not to be trusted. William L. Connor, Superintendent of Schools in Republic, Michigan, had pointed out, for example, that if the teacher were to be rated at all, it should be by results secured, not by clothing worn or methods chosen. Connor wrote, in 1920, that "It is the teacher's task to make changes for the better in the abilities, habits, and attitudes of boys and girls. Her efficiency can be evaluated fairly only in terms of her success at this task." In fact, Connor quoted, most principals had discovered that "The homely girls make the principles work, while the pretty girls work the principlas. . . . "[9]

Even so, Barr knew that those responsible for making judgments about teaching quality could not be depended upon to agree about how to evaluate what they observed. The criteria by which their judgments were made were often

---

[8]A.S. Barr, Elementary School Standards for the Improvement of Teaching (Ann Arbor, MI: Edwards Brothers, 1924), n.p.

[9]William L. Connor, "A New Method of Rating Teachers," Journal of Educational Research (May 1920): 338.

curious. As Assistant Director in Charge of Supervision for the Detroit Public Schools, Barr had discovered that fifteen principals observing the same recitation at the same time had fifteen different opinions about the quality of the teacher's work.[10]

He tackled the problem by working with his school supervisors, in Detroit, to list what made good teaching in terms of definite, observable, objective teacher and pupil growth activities that could be used to secure uniform judgment. In 1924 the Detroit supervisors had completed activity analysis for every subject. Their work was published as Elementary School Standards for the Improvement of Teaching. Barr introduced the new standards. His enthusiasm for research and for the application of scientific methods to education permeated the manuscript. Two ideas were central: "(1) greater economy of time and effort in the improvement of teaching; (2) the development of understandable, objective observable standards."[11]

Barr's Elementary School Standards is replete with lists of "definite, objective, observable" activities. An example that he referred to again in later publications was the list prepared by Julia Grant, Supervisor of Household Arts for the Detroit Schools:

---

[10]Barr, Elementary School Standards, p. 3. A few years later Barr was to ask sixty supervisors to analyze the same classroom recitations and rate the teacher on a commonly used rating scale. He had expected some disagreement, but was surprised by its extent. Some supervisors would have dismissed the teacher for the very reasons that endeared her to others. See A.S. Barr, Characteristic Differences in the Teaching Performance of Good and Poor Teachers of the Social Studies (Bloomington, IL: Public School Publishing Co., 1929), pp. 3-8.

[11]Ibid., n.p. Barr acknowledged that the title was inspired by Frank McMurray's Elementary School Standards as well as his deep respect for the work of the Committee on Economy of Time in Education.

Items to Observe in Sewing

1. Has each girl her work?
2. Does each girl wear a thimble?
3. Does each girl take a good working position?
4. Do the girls use straight needles?
5. Are the pins in at right angles to the edge?
6. Is the thread well fastened at the beginning and end of the seam?
7. Are the rows of stitches or seams straight?[12]

"Sewing, specialized and technical as it is, when analyzed is objective and simple of observation," Barr pointed out. "Has each girl her work? Either she has or she has not. That is, this time is objective, understandable and observable." Even more, the observation data could be quantified if the observer chose to focus on the girls who had and who did not have their work, did or did not wear a thimble or take a good working position. The great advantage of this kind of specifity was that, having agreed that such items constituted indices of good instruction in sewing, the judgments of different observers were comparable.[13]

Though some might argue that standards of this sort formalized instruction, Barr dismissed such ideas. Standards would deal with practices "that science and experience have established as best (right) beyond doubt." If there were several equally acceptable ways, the standards could be flexible. Barr hoped that standards would eventually serve as abbreviated lists of essential elements that could guide the teacher in his or her own self-improvement. Though admittedly less picturesque than other ways of describing teaching, Barr believed that standards were worthwhile because they were "thought saving, labor saving, and

---

[12] Ibid., p. 3.

[13] C.J. Anderson, A.S. Barr, and Maybell G. Bush, Visiting the Teacher at Work: Case Studies of Directed Teaching (New York and London: D. Appleton & Co., 1925), p. 4.

time saving."[14]

No mention is made in Elementary School Standards of aspects of instruction that could not be directly observed in a uniform, objective manner, like an experimental attitude, a sense of style, or satisfaction derived from sewing. While Barr was pleased that a technical subject like sewing could be reduced to uniform standards, it did not seem to have occurred to him that this is precisely why it could be so readily reduced to objective standards. He bypassed the injustice to subject that occurs when aesthetic and artistic subjects must be reduced to their technical elements for purposes of creating objective, uniform criteria. Indeed, Barr believed that such distortion was necessary in order that statements of criteria might be more serviceable.[15]

In Supervision of Instruction, written with William H. Burton in 1926, Barr suggested that through activity analysis, of the kind undertaken in Detroit, one could identify "small items of actually observable behavior on the part of the teacher or pupil." By referring to such lists, a supervisor could study a classroom and objectively evaluate the teacher's work. "We cannot observe the process of learning as it goes on, since it consists of changes made within the

---

[14]Barr, Elementary School Standards, pp. 7, 6.

[15]This is especially apparent in the section on music, written by Fowler Smith, Assistant Supervisor of Music. Smith described music as "a medium of expression of thoughts and feelings beyond the power of the spoken word." To enable the teacher to help students develop "scholastic respect" and "emotional regard" for music, were technical activities: "Are voices tested individually at the beginning of the term?" or "Does the teacher allow time for the pupils to listen to the pitch individually?" Non-technical aspects of music were less crisp: "Is there a sympathetic understanding between teacher and pupil?" or "Do pupils show an appreciation of the charm of rhythm as evidenced by apparent pleasure, facial expression, or comment when they are given the opportunity to give free expression of accent, or to respond with rhythmic movements or dancing?" Barr, Elementary School Standards, pp. 126-37.

learner, but we can observe certain activities which are known to connote or are supposed to connote learning or interference with learning."[16]

Barr was soon to reject the idea of activity analysis as the means to discovering the criterion of efficient teaching, in favor of measurement of the acts of teaching as manifest in the products of learning. In 1927 he was to write that some of the activities included in Supervisoion of Insruction were almost absurd in their attention to detail. "It is in the performance of these activities that real differences in teaching are found."[17] Though he moved away from advocating use of activity analysis, a theme popular with those who had applied principles of scientific management to curriculum making, Barr had strongly identified with the line of thought that had given rise to it. His early works illustrate the extent to which he considered his own efforts to be part of a significant movement toward the development of a science of education and those persons he identified as its leaders.[18]

Measuring the Products of Teaching

When Barr began to think of measuring teacher efficiency as it was manifest in the products of learning, he stepped into what was a developing strand of

---

[16]A.S. Barr and William H. Burton, The Supervision of Instruction: A General Volume (New York and London: D. Appleton & Co., 1926), pp. 117.

[17]A.S. Barr, "Some Problems Ahead in the Scientific Study of Classroom Supervision," Journal of Educational Method 7 (November 1927): 69.

[18]In the first chapter of An Introduction to the Scientific Study of Classroom Supervision, Barr connected the emergence of scientific supervision with "the history of the scientific study of education itself and with the larger movement for the scientific study of physical and biological phenomena." Included in his list of contributors to this movement are: Frances Galton (Hereditary Genius is mentioned), J. McKeen Cattell, A. Binet, Joseph Rice, E.L. Thorndike, S.L. Courtis, Franklin Bobbitt, and W.W. Charters. See A.S. Barr, An Introduction to the Scientific Study of Classroom Supervision (New York and London: D. Appleton & Co., 1931), p. 1.

research into use of educational tests and measurement. His studies at Indiana University, Chicago, and Wisconsin had brought him into contact with the newest theories of human learning--notably the theory outlined by E.L. Thorndike and expounded by William A. McCall and Charles H. Judd. Thorndike's view of education provided the underpinning for development of a science of education:

> Education is concerned with changes in human beings; a change in a difference between two conditions; each of these conditions is known to us only by the products produced by it--things made, words spoken, acts performed, and the like.[19]

If followed that if the products of education were to be measured, they must first be defined in such a way that their amount could be quantified. The difficulty inherent in doing this came from the fact that not all of the products of learning were manifest in an outward, observable product.

There had been some limited interest in uses of tests and measurement for determining suitability of teachers for their work, but only a few scholars attempted to link measurement of the products of learning to efficiency in teaching. S.A. Courtis had suggested that teaching efficiency ought to be defined wholly in terms of pupil change, challenging educators to develop tests that would measure pupil change.[20] Others, by far in the majority, looked to rating scales as a more useful measure of teaching efficiency.[21] Even William

---

[19] E.L. Thorndike, "Measurements of Educational Products," p. 14.

[20] Stuart A. Courtis, "Standards of Teaching Ability," Educational Review 62 (1921): 183-86.

[21] See, for example, Harold O. Rugg, "Self-Improvement of Teachers Through Self-Rating: A New Scale of Rating Teachers' Efficiency," Elementary School Journal 20 (1820): 670. There were three commonly used methods for evaluation of teachers: 1) rating by means of a rating scale, 2) testing teaching ability, 3) measurement of changes produced in students. In the period between 1900-1930, roughly ten percent of the articles written on teacher efficiency

A. McCall, who believed that "the only just basis for selecting and promoting teachers is the changes made in pupils," did not believe that tests had reached a state of development that warranted their application to the evaluation of teachers. He feared, in fact, lest the values of tests in the diagnosis and guidance of students be jeopardized by antagonizing teachers in attempts to measure their efficiency.[22]

Nevertheless students of McCall at Teachers College, Columbia University adopted the pupil gains criterion of teaching efficiency, attempting to deal with some of the problems of testing that limited its usefulness. Lelah Mae Crabbs, for example, devised a combination of efficiency measures in teaching all subjects, which she utilized in Queen Anne's County, Maryland. Crabbs concluded that teaching efficiency could not be judged by supervisors, by testing the teacher's knowledge, but by means of measuring with the efficiency formula.[23] Her study earned Crabbs a Doctor of Philosophy degree and was deemed worthy of publication as a Teachers College, Columbia University Contributions to Education volume, in 1925.

---

deal with measurement of pupil growth or gains as a criterion of teacher efficiency. By and large, articles were devoted to development of standards including rating scales. Other articles dealt with ability tests, merit pay, and general discussions of efficiency in teaching.

[22]McCall, How to Measure, pp. 150-51. McCall offered a technique which he believed would be useful in rating and selection of teachers. He attributes the actual formula to Franzen. It requires computing of an Accomplishement Quotient (A.Q.), or reading age divided by mental age--a series of ten steps. The Reading A.Q. Difference would be taken form the mean initial A.Q. subtracted from the mean final A.Q. N = number of subjects tested. The formula read:
Teacher Efficiency = $\frac{\text{(Reading A.Q. Diff) + (Arith. A.Q.) + ...}}{N}$

[23]Lelah Mae Crabbs, Measuring Efficiency in Supervision and Teaching. Teachers College, Columbia University Contributions to Education, No. 175 (New York: Teachers College, Columbia University, 1925).

Barr took seriously McCall's interest in and cautions about the use of tests of pupil gains in measuring teacher efficiency. But, since the school existed primarily for the child, he reasoned, "The teacher who accomplishes most toward educating the child is the most efficient. Broadly interpreted, the teacher is best who produces the greatest number of desireable changes in pupils."[24] In chapter seven of his book, Scientific Supervision, "The Measurement of the Results of Supervision," Barr paraphrased and quoted extensively from chapter five, "Measurement in Evaluating the Efficiency of Instruction," of McCall's How to Measure in Education. In doing so, he made available information on testing and measurement to the student of school supervision. Barr passed along McCall's warning, suggesting that "The use of test scores for evaluating the efficiency of teachers is an exceedingly delicate matter. All told, more harm may be done than good."[25] Objective measures were to be considered useful only to the extent that they did not "seriously disturb the more important subjective features" of teaching. Activity analysis and standardization of objective items for guiding classroom observation had not been satisfactory measures, in part because they did not account for the many products of learning that Barr believed to be important.

In Barr's view, the products of learning were to be understood broad in terms.

> The products of learning are the abilities, knowledges, skills, attitudes, and ideals acquired by the pupils taught. They may be physical, mental, social, or moral. They may pertain to the pupil's ability to maintain himself in society, his language activities, his health activities, his citizenship activities, his general social activities, his leisure time activities, his religious activities, or

---

[24] Barr, Scientific Supervision.

[25] Ibid., McCall reads, "All told more harm may easily be done than good," p. 150. The same information was disseminated through the textbooks written with William H. Burton and L.J. Brueckner.

> to his future parental and vocational activities. They include concomitant learnings as well as primary learnings.[26]

On the one hand, Barr was anxious to find objective, measurable criteria of teaching efficiency, believing that "The best measures of teaching are to be found in the kinds and amounts of changes produced in individual learners." Learning was to be compared before and after instruction, "If no change has taken place, then no learning has taken place."[27] On the other hand, he was concerned that many important aspects of learning--indeed, the most important aspects--could not be expressed in objective measurable form. "Individuals, devoted to the task of developing objective tests, are particularly prone to regard the now measureable products of learning as those worth measuring."[28] Tests of knowledge and skills, to the neglect of qualities like "Attitudes, ideals and habits of conduct," could potentially "change the course of instruction along the channels emphasized in measurement, and the repeated use of the same sorts of instruments in desireable or undesireable standardizations."[29]

Barr wrestled with the value and limitations of objective measurement of teaching efficiency, arguing both for and against the use of such measurement. His arguments appeared in articles, and, almost monthly, in editorials for the Journal of Educational Research and the Journal of Experimental Education.

---

[26] Ibid., p. 10.

[27] Barr and Burton, Supervision of Instruction, p. 294.

[28] Barr, Scientific Supervision, p. 102.

[29] A.S. Barr, "The Effects of Measurement of Instruction," Journal of Educational Research 28 (March 1935): 485. Earlier he recalled an experience in preparing a geography curriculum consonant with progressive ideals, only to have the school district issue a geography test. "It is the writer's opinion that the new test did more to determine the character of geography teaching in that school system than did the course of study." Scientific Supervision, p. 67.

He reasoned that teachers had been, and always would be, evaluated. "The fair treatment of teachers and the improvement of teaching depends upon the more accurate measurement of teaching ability."[30] Without knowing which qualities were essential to teacher success, and how to measure them, progress in the study of teaching could not be made. The measurement problem claimed more and more of Barr's attention.

He had begun with the question of how to best supervise the teacher, asking, "How do you know a good teacher when you see one?" In the late 1920s, Barr had seen the question as a supervisory problem, addressing it in his Appleton Series textbooks on supervision, An Introduction to the Scientific Study of Classroom Supervision and Supervision of Instruction: A General Volume, written with William H. Burton. By 1947, when the second edition of the general textbook on supervision was published, Barr was assuming a more limited role. He was beginning to see the original question not as a supervisory problem, but as a measurement problem that could not be left to the supervisor.

His message in the 1947 textbook was as it had been in the earlier edition. Barr relied on McCall's assumptions regarding measurement of changes in students as a criterion of teacher efficiency. He continued to describe the method as "exceedingly difficult," yet "theoretically sound." "If the precautions [outlined by McCall in 1922 and again in 1939] . . . suggested are kept in mind, the pupil-growth index of teaching efficiency holds considerable promise not only for evaluating efficiency but for the study of the relationship of various teacher

---

[30] Ibid.

factors to pupil growth."[31]

By the time the Appleton Series textbook was due for its final edition, Barr had completely bowed out. Burton and Brueckner were left with the supervisory problem. Study of the problems inherent in finding a criterion of teaching efficiency was Barr's consuming interest. He devoted attention to ascertaining "the competencies, qualities and behavior controls essential to success in teaching" in order to evaluate practice relating to identification of the prerequisites to teaching success and validity and reliability of those instruments used to evaluate and predict teaching efficiency."[32]

The Measurement Problem

Thorndike had spoken of the need for study of problems of measurement for measurement's sake in order to advance development of a quantitative science in education.

> The danger is that the attention of investigators will be distracted from the problems of pure measurement for measurement's sake, which are a chief source of progress in measuring anything. Perhaps not even one person in a million need feel this passion, but for that one to cherish it and serve it as far more important than for him to devise a test which thousands of teachers will employ.[33]

In a paper delivered at the Denver meeting of the Department of Educational Research at the National Education Association convention of 1935, Barr's growing

---

[31] A.S. Barr, William H. Burton, and L.J. Brueckner, Supervision: Democratic Leadership in the Improvement of Learning. 2d ed. (New York and London: D. Appleton-Century Co., 1947), pp. 379-80.

[32] A.S. Barr, "The Measurement and Prediction of Teaching Efficiency: A Summary of Investigations," Journal of Experimental Education 16 (June 1948): 203.

[33] Thorndike, "Measurements of Educational Products," p. 20.

passion for solving the measurement problem related to the criteria of efficient teaching was made clear. He called for development of functional tests that would measure the teacher in action, pointing to shortcomings in approaches to measurement of teacher ability. "I think that we do frequently overlook the fact that the teacher brings into a given teaching situation not merely a knowledge of subject matter taught or method of teaching but his whole active self."[34] Considerable time and effort had been wasted in attempts to measure teaching ability because researchers had failed to recognize that "technics" of teaching were "numerous, fleeting, transitory" acts, the observable superficial aspects of teaching.

> In the first place I suppose it goes without saying that it will be necessary to give more attention to the development of the criteria of teaching success. This is, of course, a bit embarrassing inasmuch as there is so little agreement among educationalists as to what good teaching is. It seems to me, however, that the ultimate criterion of teaching success will have to be found in the changes produced in pupils measured in terms of the objectives of education. And I think it also goes without saying that these changes or products of instruction will have to be considered broadly. We shall have to take into consideration not merely knowledges and skills and the more tangible outcomes of instruction, but attitudes and ideals and the less tangible outcomes of instruction. This in itself presents a very difficult problem for the expert in measurement.[35]

Barr's personal research agenda was set. With a commitment that would likely have placed him with Thorndike's "one in a million," Barr studied the measurement problem carefully and comprehensively.

Articles and editorials dealt with aspects of the measurment problem. In 1940, for example, he spoke of four approaches to measurement of pupil growth

---

[34] A.S. Barr, "The Measurement of Teaching Ability," in Proceedings of the Seventy-third Annual Meeting of the National Education Association (Washington, D.C., 1935), p. 339.

[35] Ibid., p. 341.

and achievement: mental prerequisites, overt behavior, qualities of the personal, and measured or inferred effects. Overt behavior was the newer approach, its shortcomings were less well known, and it illustrated the complexity of measurement. Barr devoted most of the article to a critique of this approach. The difficulty, he pointed out, was that one could make inferences about products of learning from observable pupil behavior, but acts in themselves were neither good nor bad.

> Suppose that one observes some forms of individual behavior and wishes to pass judgment on it. How does one know in the first place that the behavior is not faked, and is not an honest repesentation of the individual hopes, beliefs, and ambitions? But passing this by, how does one know that the behavior observed was appropriate to the situation giving rise to it?[36]

The tools available to the researcher concerned Barr as well. He wondered if statistical procedures were appropriate to complex problems of the teacher in the field. He wrote, in 1945, "The processes of mathematics as most of us know them have been developed for the study of physical phenomena; we sorely need a Newton in the mathematics of human relationships, individual growth and learning."[37] Procedures useful in the sciences needed to be reexamined as to their applicability to educational problems. When facts were isolated for study, they were removed from the context that gave them meaning.

Yet, it would be through solving the problems related to measurement that advances would be made in education. "Measurement is fundamental to all teaching, and professional advancement. Effort expended in this field should bring both

---

[36] A.S. Barr, "The Growing Complexity of Measurement," Journal of Educational Research 33 (May 1940): 711.

[37] A.S. Barr, "A Look Ahead," Journal of Educational Research 38 (January 1945): 381-82.

personal satisfaction and professional recognition."[38]

Ten years after his remarks before the Department of Educational Research, in Denver, Barr wrote again:

> We very obviously need more reliable means of evaluating our programs of action; whether current statistical procedures are adequate and appropriate or not for these purposes needs to be ascertained. Very frequently the assumptions upon which these procedures are based are violated in educational research. Persons not well acquainted with mathematical lore fail to note that most procedures can be used only under a very restricted set of conditions.

The criteria of efficient teaching had not been developed. The difficulty was not with mathematics, in fact, less objective means would not do. The difficulty came from the often questionable use of mathematics. To Barr, the challenge was to invent "objective tools designed for education."[39]

The Effective Teacher

As Barr continued to study measurement of teaching efficiency and to call for broadly conceived approaches to the products of teaching, a subtle shift occurred in the language that he used to describe his search. It was a shift reflected in the wider literature on evaluation of teaching, one of which he was aware. He wrote, in 1947:

> We are ultimately concerned with the teacher in action, and our ultimate goal is to improve this action. We have referred to this goal variously in broad terms, as the modification of teacher behavior, as the development of efficiency in performance, and as the improvement of teaching methods.[40]

---

[38] Barr, "Planning an Active School Year," Journal of Educational Research 31 (September 1937): 50.

[39] Barr, "Measurement and Prediction of Teaching Efficiency," p. 203.

[40] Barr, Burton, and Brueckner, Supervision: Democratic Leadership, p. 620.

Though the language of scientific management was often being set aside for other terms, like teacher effectiveness, the goals were not. The question of how to objectively measure the value of the teacher's work remained. Interestingly enough, the measure most compatible with a technological, systems model for curriculum and teaching was the one that began to receive most favor: measurement of pupil growth and achievement as manifest in observable behavior or on standardized tests.

Barr selected the pupil gains criterion for use in the Wisconsin Studies of Teaching Ability, begun in 1940, in cooperation with doctoral students at the University of Wisconsin. He accepted that it was theoretically sound, but continued to admit that even so, "its use presents many very real difficulties." An attempt was made to measure changes in pupil attitudes, skills, and behaviors, as well as in the acquisition of information, before and after teaching.[41]

At the same time, Barr still believed that "no wholly satisfactory criterion has been developed" to measure teacher effectiveness. Without more thorough investigation "there is probably no justification for use of one approach to the exclusion of others."[42]

Barr was not alone in his search for the criterion of teacher effectiveness. In 1950 he was appointed to serve on the AERA Committee on the Criteria of Teacher Effectiveness. The group was challenged to define problems of effective teaching. A conceptual framework that could be used in planning studies of teacher effectiveness was developed. "In identifying effectiveness as a criterion

---

[41] A.S. Barr, "Impressions, Trends and Further Research," Journal of Educational Research 14 (December 1945): 200.

[42] A.S. Barr, "The Use of Measurement in the Management of Teacher Personnel," Education 66 (March 1946): 435.

dimension of teachers, we imply that the purpose of our measurement or appraisal of teachers is to estimate whether they will produce desired amounts and types of changes in pupil behavior."[43] To have an effect, the teachers behaved in some ways. "Thus, if we think of our various types of phenomena as serially linked, with 'effects' as phenomena at one need of the chain, then the behaviors of teachers may be considered as phenomena that occur before the effects," the committee reported. Hypotheses relating behaviors and effects were needed. "On the wisdom of our solution to this question probably depends the realization of a science of teaching and a scietific basis for such programs as teacher selection, training and supervision."[44]

These were concerns that Barr had aired over the years. In the same issue of Review of Educational Research, he wrote:

> Teaching effectiveness generally has been treated as something apart from the situations giving rise to it. More needs to be known about the situational determiners of effective teaching.
> . . . Much of the research on teacher effectiveness seems to proceed as if the qualities in question resided entirely in the teacher. This may or may not be true. "Teacher effectiveness" may be essentially a relationship between teachers, pupils, and the other persons concerned with the educational undertaking all affected by the limiting and facilitating aspects of the immediate situation.[45]

Barr had found the pupil and teacher behavior approach interesting and promising. He adopted pupil gains as the criterion for the Wisconsin Studies. Yet he had never accepted an approach without acknowledging its limitations, dealing with

---

[43] A.S. Barr et al., "Report of the Committee on the Criteria of Teacher Effectiveness," Review of Educational Research 22 (June 1952): 243.

[44] Ibid., pp. 259-60.

[45] A.S. Barr, "The Measurement of Teacher Characteristics and Prediction of Teacher Efficiency," Review of Educationl Research 22 (June 1952): 172.

alternatives, and calling for a broader look at the teacher and pupil than could be seen in narrow definitions of the products of learning. Whether or not Barr greatly influenced the Committee on Criteria of Teacher Effectiveness, its reports related to the concerns that he had raised for three decades.

The second report of the Committee on the Criteria of Teacher Effectiveness was made in May of 1953. The committee had done little more than call for a theoretical framework for guiding research and critique the direction that research had taken in the past. "In the earlier decades of the century, research seems to have been aimed at discovering the characteristics of the effective teacher."[46] Research had too often proceeded "without explicit theoretical framework, in intellectual disarray, too often to the testing of myriads of arbitrary, unrationalized hypotheses." Teaching must be seen as a complex activity. For all of its effort, the committee concluded that after examining forty years of research on teacher effectiveness, it could see no evidence that would serve to guide any one needing to make decisions regarding the training of teachers, pre-service or inservice, or personnel decisions relating to teachers.[47]

Barr continued to question links between observed teacher behavior and pupil gains as measured by objective instruments. Yet he continued to look at both teacher and pupil behavior. Sooner or later one would need to assign values to behaviors in order to judge them--this was the troublesome side of

---

[46] A.S. Barr et al., "Second Report of the Committee on the Criteria of Teacher Effectiveness," Journal of Educational Research 36 (May 1953): 643-44. Members of the commitee were Barr, Burley V. Bechdolt, Warren W. Coxe, N.L. Gage, Jacob S. Orleans, H.H. Remmers and David G. Ryans. The committee was not always in agreement, and Coxe filed a "minority report" to the 1952 document and resigned from the committee, believing that it should have devoted attention to coming up with actual criteria rather than principles.

[47] Ibid., p. 657.

studying behavior. "Nothing can be more misleading than counting behaviors out of context," he warned. "Where feelings, knowledges, and skills are used to measure teacher competency, one runs into the danger of remoteness. These prerequisites to competency do not function in a vacuum. They are or should be transformed by the demands of the situation."[48]

"How do you know a good teacher when you see one?" To know, one would need to be able to identify the essence of teaching, its indispensable essentials. By 1960, Barr was convinced that it was to be found through observation of behavior, but that it was part of a complex relationship. Pupil growth or achievement always found its way back to the performance of the teacher, whose behavior was subject to observation. Research seemed to support the notion that teacher personality was important to teacher effectiveness; this being the case, theories of personality could relate to traits of persons. "Another way of looking at the subject, however, is to consider not qualities of the person but characteristics of performance or behavior." Both were elements of personality, characteristics of performance having the advantage of being more subject to observation.[49]

Any approach would have limitations. Hence, accepting a behavioral approach, with all of its limitations, was as reasonable as any other. It was, perhaps, more so to Barr, since a behavioral approach had the advantage of making it possible to integrate the concept of personality with methods of teaching.

---

[48] A.S. Barr, "Problems Associated with the Measurement and Prediction of Teacher Success," Journal of Educational Research 51 (May 1958): 696.

[49] A.S. Barr, "The Assessment of the Teacher's Personality," School Review 68 (Winter 1960): 400.

> Psychology has defined the conditions for effective learning in terms of certain principles of learning. This suggests the question: Are the techniques of teaching that are presumed to grow out of learning theory encompassed by the behavioral aspects of personality, or are the behaviors found in the techniques of teaching something different? Possibly the techniques of teaching grow out of another constellation of values that one needs to keep in mind. Possibly there is some unique constellation of human relationships that behavioristically constitute the essence of personality and another constellation of teaching activities that constitute technical competency.[50]

Barr acknowledged that he was being pragmatic in deciding to stay with a behavioral criterion. Yet, he still wondered if one could define the work of a reliable, emotionally stable, and resourceful teacher in behavioral terms. All terms needed operational definition if one accepted a behavioral criterion, but did such defintions allow for important subtle meanings? And could observance of a behavior be considered as presence or absence of a particular characteristic?[51]

The Wisconsin Studies of Teacher Effectiveness were completed in 1961. In reporting on the studies, Barr expressed the hope that with the development of new techniques it would be possible to extend observability to internal behaviors as well as to external behaviors in the future. As to approaches to criterion development, each of the four main approaches most commonly used had advantages and disadvantages--a point he had made before.

> Each presents very real problems from the point of view of developing valid and reliable criteria. For the time being it might be best to look upon the several criteria as complementary rather than antagonistic. The strengths can possibly be preserved and the weaknesses avoided by a discrete choice from among them according to the demands of the situation.[52]

---

[50] Ibid., p. 401.

[51] Ibid., p. 402.

[52] A.S. Barr et al., Wisconsin Studies of the Measurement and Prediction of Teacher Effectiveness: A Summary of Investigations (Madison: Debar Publications, 1961), pp. 11-12.

Barr's conclusion to the summary of the Wisconsin Studies was that progress had been made in clarifying the problem, but much more work was needed to "clear up many unsolved problems relative to the measurement and prediction of teacher effectiveness."

"How do you know a good teacher when you see one?" The question remained after forty years of effort to develop a more objective criterion of teacher effectiveness. Barr had not solved it to his satisfaction.

Conclusion

When A.S. Barr worked with teachers and supervisors in the Detroit Public Schools, he was concerned with classroom supervision. His initial efforts were devoted to making the supervisory process less capricious. He soon became convinced, however, that supervision of instruction could not be improved without some uniform understanding of what constituted good teaching. He set out to establish objective, valid criteria of efficient teaching, seeing his own efforts as part of a greater movement toward development of a science of education. This effort led him further away from supervision and into the study of the problems related to measurement of teacher efficiency. As the times changed, his language changed. The search was described as a search for the criteria of effective teaching. Yet the goals of the search remained constant.

By adopting pupil growth and achievement as the criteria of effective teaching in the Wisconsin Studies, Barr endorsed measurement of pupil growth and achievement as the criteria of effective teaching whether or not he intended to do so. While he respected the limitations of the pupil growth and achievement criteria, it is not certain that those who were searching for greater precision and control in curriculum planning and implementation understood

these limitations. When the Wisconsin Studies were finally summarized for publication, public school people were looking for ways to prove their effectiveness. In the early sixties, schools were still recoiling from the nearly devastating public attacks by individuals like Arthur Bestor, publicized on the pages of Life and U.S. News and World Report. These attacks had questioned the very purposes of schooling, calling for greater attention to the development of intellectual abilities and of skills. And, in the aftermath of Sputnik, Americans were calling, even more persistently, for attention to intellectual goals of education and for massive infusions of mathematics, science, and technology into the curriculum. Hence, it may not have been af as great a concern to some educators as it was to Barr, that pupil gains criteria focused most on knowledge and skills, that could be measured by objective means, at the expense of other "products" of learning that could not.

There is no evidence that it was A.S. Barr who, in the final analysis, furnished the basis for objective-based criteria of teacher effectiveness as measured by pupil growth and achievement. It is certain, however, that he wrote and published more on the topic than any other researcher during the period from 1900 until 1950.[53] And, while he was cited frequently by those interested in studying teacher efficiency, effectiveness, or competency (the 1970s update of teacher efficiency), Barr's contribution was probably more in keeping with the general flow of events than it was trend-setting. Yet it is significant to note that apparently the message heard from Barr was not about advantages and limitations of the pupil gains criteria, but about its advantages. For example,

---

[53]Simeon J. Domas and David V. Teidman, "Teacher Competence: An Annotated Bibliography," Journal of Experimental Education 19 (December 1950): 101-208. Twenty-five of 672 annotated references are by Barr.

when Simeon J. Domas and David V. Teidman, of the Harvard Graduate School of Education, compiled their comprehensive bibliography on teacher competence, covering fifty years of research, every attempt was made to annotate objectively and accurately. In annotating Barr's "Measurement of Teaching Ability," taken from his 1935 address to the Educational Research Association, Domas and Tiedman wrote:

> Measurement of teaching ability is necessary for the institutional, and in-service training of teachers, the administration of the teaching staff, and research in tehse areas. Shortcomings in approaches to the measuremt of teaching ability and the need for development of functional tests measuring the teachers in action were discussed. Certain criteria of teraching success were mentioned. Changes produced in pupils, measured in terms of the objectives of education, is the ultimate criterion of teaching success.[54]

The description is accurate, yet placement of the final sentence highlights the pupil gains criterion and removes the cautions that Barr placed around it.

Another example of the emphasis given to Barr's work is to be found in an article by James W. Trent, Arthur M. Cohen, and Clare Rose, in the Second Handbook of Research on Teaching. Barr is quoted as having concluded that pupil growth or achievement is "the primary criteria against which all other criteria should be validated."[55] Cohen's own argument had been that student achievement of learning objectives is the main criterion on which studies of faculty and instructional efforts should be based."[56] The quotation cited does not

---

[54]Ibid., p. 111; italics mine. Refer to page 14 of this paper for the context in which Barr's comment was placed.

[55]Trent, Cohen, and Rose, "Research on Teaching in Higher Education," p. 1049; italics mine.

[56]Ibid. (See A.M. Cohen and F.B. Brawer, Measuring Faculty Performance. ERIC Clearinghouse for Junior Colleges, Monograph No. 4 (Washington, D.C.: American Association of Junior Colleges, 1969.). ED 031-222.

come from the conclusion to the Wisconsin Studies, nor is it offered as a conclusion. It is to be found in Barr's introduction and it reads in full:

> A fourth type of criterion of teacher effectiveness is that of pupil growth and achievement, which is usually expressed as pupil gain scores based upon achievement tests administered prior to instruction and again at some subsequent date when a particular unit of instruction or course has been completed. To most persons this criterion is considered a primary criterion against which all other criteria should be validated. Like the other criteria it is subject, however, to very definite limitations.[57]

This accident of research may illustrate not so much the kind of editorial slip to which all researchers are subject, as it illustrates the way that Barr's work was received. Barr did not conclude that the primary criterion of teacher effectiveness is pupil growth, but a primary criterion to most persons. Yet it is clear that at least to Trent, Cohen, and Rose, Barr's work supported their own view. It is likely that Barr's message was received, not only in the research community, but by the public as confirmation that the teacher's worth should be measured by what the teacher is able to produce in the way of student outcomes on objective measures of growth and achievement.

Development of the criteria of teacher effectiveness may not appear to be a curriculum concern. Curriculum has more to do with theories and processes of decision making about what should be taught, to whom it should be taught, and how. Teacher effectiveness seems more related to professional evaluation. But as the teacher's role in controlling inputs into educational systems has become more specifically defined, it has been possible to define points at which the teacher is accountable for educational outcomes as a measure of the teacher's competency. Not every teacher is content to be involved in curriculum at the

---

[57] Barr et al., Wisconsin Studies, p. 13; italics mine.

implementation level, seeing the role of the teacher as that of curriculum leader in the classroom. Yet the way that the teacher actually becomes involved in curriculum design and decision making is often defined by circumstances other than his or her preferences, style, or beliefs.

When the teacher knows that his or her worth will be the topic of discussion once the school test scores are published in the local paper, the teacher is advised to make preparation of students to take the tests a curriculum concern, school policy and personal preference notwithstanding. When school administrators can point to general or composite measures of pupil gains as evidence of the teacher's competence, no one need tell the astute teacher which curriculum design model to follow.

Those who still puzzle over the persistence of the technological systems approach to curriculum theory and practice may wish to take note.